

動画に対するボイスコメントの合成再生システムの提案

Proposal for a system for synthesizing and replaying voice comments on videos

田村和也[†] 川合康央[†]
Kazuya Tamura[†] Yasuo Kawai[†]

[†] 文教大学 情報学部

[†] Faculty of Information and Communications, Bunkyo University.

要旨

本研究は、音楽ライブなどの映像コンテンツに、ユーザの掛け声や歓声などのボイスコメントを追加し、音声情報を合成・再生するシステムを提案するものである。試作したシステムでは、音声コメントの重なりやノイズ、合成される音域の大きさなどにより、聞き取りにくいケースがあることが明らかとなった。そこで本稿では、ボイスコメントを加工処理することによって、音声情報の明瞭度を向上させることができるかについて検討を行うこととする。

1. はじめに

動画投稿サービスは、多くの人々に親しまれているコンテンツであり、今後も需要が見込める Web サービスである。動画投稿サービスの例として、YouTube や TikTok 等のコンテンツが挙げられる。また、動画投稿サイトであるニコニコ動画では、画面上の動画にユーザのコメントが重畳表示されるインタフェースを提案している。これまでに、本研究では、ユーザコメントを文字ではなく音声として投稿できる新たな動画投稿サービスの検討と、その可能性について、試作システムの開発を通じて検証を行ってきた。本稿は、試作システムの評価の際に課題として挙げられた音質の改善についての報告を行う。

2. 先行研究

動画視聴時に、ユーザの音声を扱うことについて、松長ら[1]は、WakWakTube のユーザアバターに音声情報を加える実験を行っており、それが効果的に作用していることを示している。また、高野ら[2]は、他ユーザと共に TV を視聴しているような臨場感を持つ動画視聴システムを提案しており、会話が成り立つ精度での再生同期制御を実現した。本研究では、これら先行研究を踏まえ、独自のボイスコメントの合成再生システムの開発を行う。

3. システム開発

3.1. 開発環境

今回作成したライブ再生サイトの開発では、HTML, PHP, JavaScript を使用した。HTML でインタフェースを、PHP でサーバー側の処理を、JavaScript でクライアント側の処理を行った。開発環境は VisualStudioCode, XAMPP (Apache), chrome を用いた (図 1)。

3.2. 動画に対する音声コメントの取得方法

音声コメントは、クライアントが PC に設定しているマイクデバイスから取得するものとした。入力された音声は、Web ページに埋め込まれた JavaScript によって処理され、音声ファイルとして出力される。この時、音声ファイルには動画のどのタイミングでコメントされたものなのか、またユーザを判別するための識別番号が記録されているものとした。クライアントは、これらの情報を投稿することができ、その場合の音声ファイルは PHP によってサーバー上に記録される (図 2)。

3.3. 動画に対する音声コメントの付与方法

動画再生中には毎時、現在の再生時間を JavaScript にて取得する。もしその時間に対応した音声コメントが投稿

されている場合、そのコメントを再生する。コメントの音量はクライアントが調節できるものとしたが、デフォルトでは元の動画の音量の半分ほどに設定した（図2）。

3.4. ライブ配信に対する音声コメントの取得方法

音楽イベントのライブ配信の場合、リアルタイムに音声伝わらず臨場感を味わえないと考えられる。そのため、本システムでは WebRTC を使って、グループ通話に近い形で機能を用意することとした。今回は、Skyway という API を利用してこの機能を実装した。



図1 システムのスクリーンショット

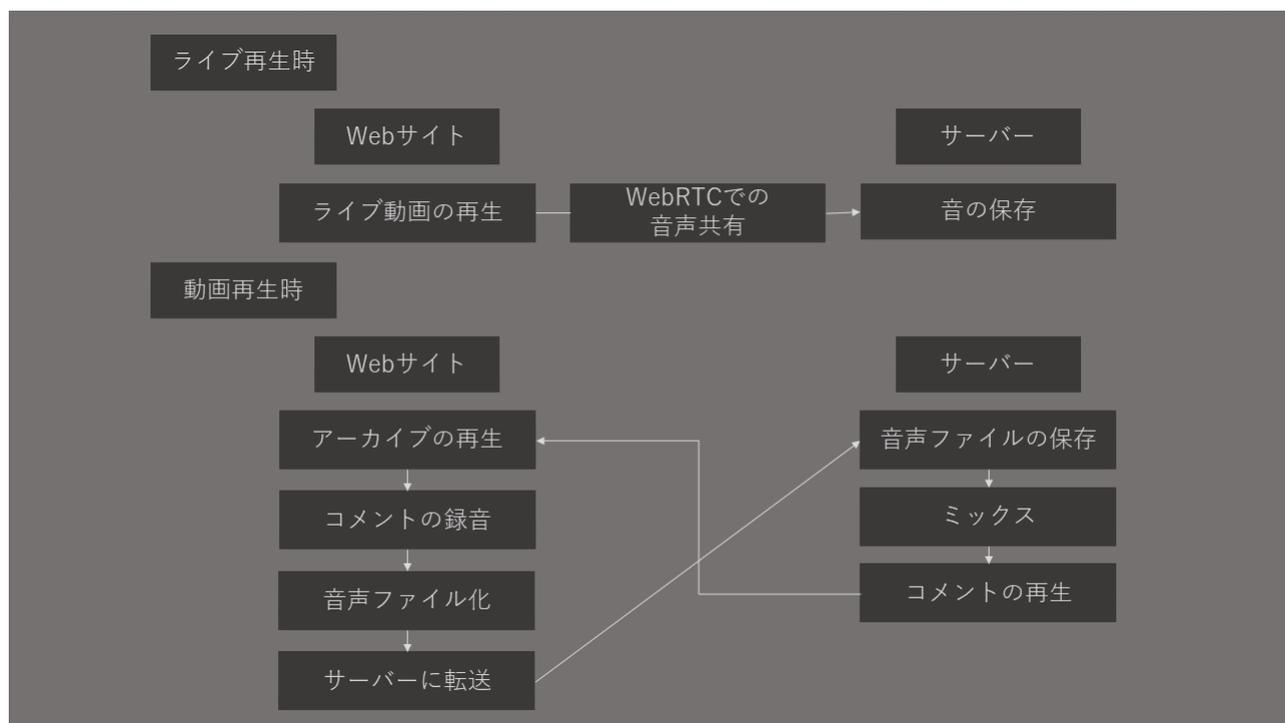


図2 システムの構成図

4. 実験と考察

4.1. 音声処理の改善

試作したシステムは、ボイスコメントを重ね続けていくと、元の音声情報が薄くなり、雑音と化していった。この問題は、ボイスコメントの重なりや雑音、合成された音域の大小によって、聞き取りが困難になっているのだと仮定し、修復実験を行った。結果、手動でノイズとなる要素を取り除いた結果、音質に改善されたため、これらが原因の一部だと確認された。

まず、雑音について、詳細に見ていくこととする。雑音については、一定以下の大きさの音を取り除くことで改善される。だが、この雑音除去処理を自動で行う場合、コメントごとに雑音を検出しなければならないため、その判別は非常に難しい。AIを用いた判別処理も検討されたが、雑音の種類は多岐にわたるため、本システムでは単純な工程を行うこととした。元となる動画の音声とは異なり、ボイスコメント自体に音質はそこまで求められないと考え、本研究では、暫定的にすべての音量の1/4の大きさをカットし、3/4を引き延ばして雑音の低減を目指した。1/4という比率は、いくつかの音源をもとに設定された暫定的な数値であるが、今後被験者実験を行い、最適な値を明らかにしていくこととする。

次に、ボイスコメントの音量を調整して合成を行った。この処理によって、コメント間の音量の差異が軽減され、聞き取りやすさが向上する。ただし、元となるボイスコメントの音量が小さい場合、引き延ばした際に音が歪んでしまうことがあるため、今後対応策を検討したい。

最後に音域の問題である。動画の音域とコメントの音域が重なると、聞き取りが困難になることが分かった。この問題に対応するため、すべてのコメントの音域を、元の動画の音域よりも音程を高くするのがよいと考えた。これは、人工的な音声感が強くなるが、実際の肉声よりも匿名性が高くなる点で、本システムにおいては適切である可能性がある。さらに、各コメントの音程を適度に散りばめれば、コメントごとの聞き分けも可能となると考えられる。今後も、引き続き評価を行っていくこととする。

5. まとめ

本研究は、動画につけられた音声コメントを合成して再生するシステムの提案を行ったものである。結果、実験

を通して、いくつかの課題とその改善点を提案することができた。一方で、ノイズの値など、粗い点が多く残っているため、システムの精度を向上させる必要がある。また、より適切な音声データの混合方法も工夫する必要がある。これらの課題に対して、今後も被験者を用いた評価実験を繰り返し、改善していきたいと考えている。

謝辞

本研究は JSPS 科研費 JP 19K12665 の助成を受けたものです。

参考文献

- [1] 松長雄也, 谷中俊介, 坂内祐一, “動画視聴における他視聴者の音声情報の再生方式”, マルチメディア, 分散協調とモバイルシンポジウム 2018 論文集, 2018, pp.530-533.
- [2] 高野祐太郎, 大島浩太, 田島孝治, 高田治, 寺田松昭, “投稿型動画視聴におけるユーザ間リアルタイムコミュニケーション支援システム”, 電子情報通信学会論文誌 D, Vol.93 No.10, 2010, pp.2302-2316.